



UNIVERSIDADE  
LUSÓFONA

# Data Science em Gastronomia Computacional

## Trabalho Final de curso

Relatório Final

Ricardo Melo

Tiago Simas

Trabalho Final de Curso | LEI | 25/06/2021

[www.ulusofona.pt](http://www.ulusofona.pt)

## **Direitos de cópia**

*Data Science em Gastronomia Computacional*, Copyright de Ricardo Melo, ULHT.

A Escola de Comunicação, Arquitectura, Artes e Tecnologias da Informação (ECATI) e a Universidade Lusófona de Humanidades e Tecnologias (ULHT) têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

---

## Resumo

Com o avanço recente da tecnologia, surge um novo campo científico denominado por gastronomia computacional que resulta da combinação das ciências informáticas com o big-data em gastronomia. A gastronomia computacional vem assim, contribuir para o desenvolvimento e expansão do conhecimento gastronómico abrindo novas possibilidades para a inovação, como por exemplo o estudo num enquadramento mundial da hipótese do food-pairing assim como, a formulação de novas hipóteses “data-driven” em gastronomia, tais como o food-bridging.

Neste trabalho final de curso temos como objetivos utilizar um dataset resultante da abstração de livros de receitas tradicionais da cozinha portuguesa que nos permite: (1) analisar o enquadramento da gastronomia portuguesa relativamente às hipóteses de food-pairing e food-bridging e (2) analisar como variam estas variáveis para as diversas regiões Portuguesas. Este enquadramento da gastronomia Portuguesa até à data ainda permanecia uma pergunta científica aberta, à qual respondemos neste trabalho.

## **Abstract**

Computational Gastronomy is a new scientific field that results from the recent technological advances, which combines computer science with big-data in gastronomy. As of today, computational gastronomy has been expanding our understanding about gastronomy not previously possible, for example, the world-wide study of food-pairing hypothesis and the data-driven formulation of new hypothesis such as the food-bridging.

In this work we have as main goals to use a dataset resulting from the abstraction of traditional Portuguese cookbooks, which will allow us to: (1) study how the Portuguese gastronomy fits into the hypotheses of food-pairing and food-bridging and (2) also study how these variables vary within different Portuguese regions. As of today, framing the Portuguese gastronomy on these important variables was still an open scientific question, which we address in this work.

---

# Índice

Resumo.....	iii
Abstract .....	iv
Índice .....	v
Lista de Figuras .....	vi
1    Identificação do Problema .....	1
2    Viabilidade e Pertinência.....	13
3    Solução Desenvolvida.....	14
4    Benchmarking.....	15
5    Resultados .....	16
6    Método e planeamento .....	26
7    Conclusão .....	27
Bibliografia .....	29
Glossário.....	30

## **Lista de Figuras**

Figura 1 – Rede de sabores	3
Figura 2 - Arestas semi-métricas e exemplos de caminhos na rede de sabores	4
Figura 3 - Média food-pairing e food-bridging	5
Figura 4 - Food-pairing versus food-bridging	6
Figura 5 - Distribuição geográfica das quatro grandes classes culinárias	7
Figura 6 - Províncias consideradas na construção do dataset.	8
Figura 7 - Fórmula para o cálculo do valor de parecença (LI - likeness value).	9
Figura 8 - Média food-pairing e food-bridging	16
Figura 9 - Food-pairing versus food-bridging	17
Figura 10 - Média food-pairing e food-bridging	18
Figura 11 - Food-pairing versus food-bridging	19
Figura 12 - Ingredientes mais utilizados por região	20

# 1 Identificação do Problema

## 1.1. Introdução:

A gastronomia computacional é um campo científico que alia as Ciências Informáticas e o Big-Data em gastronomia.

O big-data em gastronomia computacional consiste na digitalização e compilação de diversas fontes díspares de informação gastronômica. Essa digitalização provém normalmente de livros: que compilam receitas gastronômicas, afinidades entre ingredientes, e outras informações tais como a partilha de compostos químicos entre ingredientes (sabores), e outras fontes tais como a compilação de receitas em web sites, blogs etc.

Todo esse conjunto de informação digitalizada (big-data) serve de base para a aplicação de técnicas normalmente provenientes da Ciência de Dados, que vêm contribuir para o desenvolvimento e expansão do conhecimento em gastronomia, abrindo novas possibilidades com o avanço da tecnologia. Esta nova área científica denomina-se por gastronomia computacional.

A gastronomia computacional tem a potencialidade de ser aplicada a áreas tais como: (1) a impressão 3D de comida a partir de compostos base que quando combinados ajudam não só a Alta Cozinha a criar novos sabores, texturas, formas, mas também no futuro, a portabilidade de comida, e.g. a nível da possível expansão espacial. (2) a ajuda no processo criativo de um chefe de cozinha, podendo sugerir de uma forma rápida e interativa novas possibilidades de afinidades entre ingredientes. (3) a geração automatizada de novas receitas por aprendizagem automatizada. (4) na saúde, adaptando dietas específicas e personalizadas com uma maior diversidade de receitas.

O Turismo gastronômico é uma outra possibilidade onde a gastronomia computacional pode ter um papel relevante. Em geral, em gastronomia a inovação é fundamental. No Turismo gastronômico o levantamento histórico de particularidades gastronômicas regionais e sua digitalização aliada à ciência em gastronomia computacional poderão potenciar bastante este ramo. Entre muitas outras possibilidades.

O objetivo deste trabalho é enquadrar a gastronomia Portuguesa como um todo e em particular por região, no enquadramento mundial com base em dois princípios chave da gastronomia: o food-pairing e o food-bridging, previamente estudados num contexto global nos trabalhos [2,4] com base em dados gastronómicos Portugueses coletados e publicados em [5,7].

Existem duas hipóteses base que sustentam a forma como combinamos ingredientes na cozinha tradicional (world-wide) e fundamentam a gastronomia computacional: (1) food-pairing e (2) food-bridging.

A primeira hipótese introduzida pelo chef Blumenthal [1] denominada por food-pairing, consiste na ideia de que se dois ingredientes que partilhem um maior número de componentes de sabor (representam um conjunto de sensações incluindo odores -

estimulação do olfato, gostos - estimulação do paladar e frescura) têm maior probabilidade de combinarem melhor entre si do que ingredientes que não partilhem estas afinidades químicas (sabores).

A segunda hipótese é o food-bridging proposta por Simas et al. [2] que assume que se dois ingredientes não partilharem uma afinidade molecular ou empírica como descrito no food-pairing, eles podem tornar-se afins através de uma cadeia de afinidades que envolve outros ingredientes que fazem a ponte entre os ingredientes de pouca ou nenhuma afinidade.

Estes dois princípios foram analisados e estudados world-wide com base numa compilação digitalizada de mais de 56.498 receitas e a digitalização de um livro (Fenaroli's Handbook of Flavor Ingredients) [3] que relaciona ingredientes e os seus compostos químicos que se denomina por sabores (flavors em inglês).

Na base destas análises está um modelo gráfico, the Flavor Network [4], como mostra a figura 1 (A) e (B).

Como podemos observar nesta figura 1, um modelo gráfico é uma abstração matemática que pode ser manipulada matematicamente, e que consiste num conjunto de nós (nodes em inglês) interligados por traços (links em inglês). Neste caso a Flavor Network figura 1 (B), é construída a partir de uma relação gráfica bipartida figura 1 (A), em que neste último caso, os nós são representados pelo conjunto de união de ingredientes e sabores (flavors) e a sua interligação (traços entre ingrediente e sabor) representa que o par (ingrediente, sabor) existe, e quando se junta toda esta informação a partir do grafo bipartido figura 1 (A), obtemos como resultado um grafo ou rede cuja representação gráfica representa a relação entre ingredientes (nós), onde os traços nesta última relação ou grafo representam o número de partilha de sabores entre cada par de ingredientes, Figura 1 (B).

Neste contexto de Flavor Network, de acordo com [2,4], o Food-pairing é tanto mais intenso quanto maior for o valor da ligação direta (traço) entre dois ingredientes nesta rede de sabores (Flavor Network) e o Food-bridging é então a capacidade de ligar um par de ingredientes que podem ou não ter uma ligação direta, através de um caminho constituído por ingredientes não repetitivos dentro de uma rede de afinidades entre ingredientes, neste caso a rede de sabores. Em suma, em gastronomia, o Food-pairing intensifica o sabor de uma receita misturando ingredientes com semelhanças na sua composição química e Food-bridging suaviza o contraste entre ingredientes.

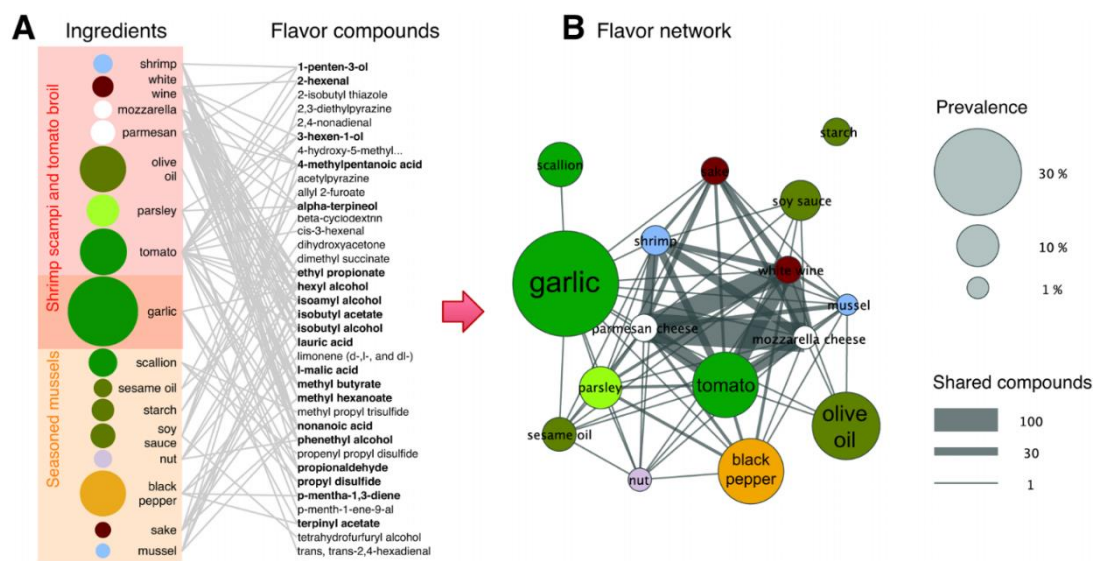
## 1.2. A Hipótese de Food-pairing:

A hipótese de food-pairing foi analisada em [4]. Aqui os autores introduzem uma abordagem em rede para explorar o impacto dos componentes de sabor nas combinações de ingredientes. Cada ingrediente foi em média ligado a 51 componentes de sabor e foram utilizadas no total 56.498 receitas.

Foi construída uma rede bipartida que consiste em dois tipos diferentes de nós: (i) 381 ingredientes utilizados em receitas por todo o mundo e, (ii) 1.021 componentes que são conhecidos por contribuir para o sabor de cada um dos ingredientes. A projeção desta rede



bipartida resulta assim na rede de sabores em que dois nós (ingredientes) estão ligados se partilharem pelo menos um componente de sabor.



**Figura 1 - A) Ingredientes contidos em duas receitas (coluna da esquerda), juntamente com os componentes de sabor presentes nos ingredientes (coluna da direita). Cada componente de sabor está ligado aos ingredientes que o contêm, formando uma rede bipartida. Alguns componentes (a negrito) são partilhados por múltiplos ingredientes. B) Projeção da rede bipartida no espaço de ingredientes, formando a rede de sabores, cujos nós são ingredientes, ligados se partilharem pelo menos um componente de sabor. A espessura das ligações representa o número de componentes de sabor que dois ingredientes partilham e o tamanho de cada círculo corresponde à prevalência dos ingredientes nas receitas.**

Conclui-se que a cozinha Norte Americana e da Europa Ocidental exibem uma tendência significativa para receitas cujos ingredientes partilham componentes de sabor. Pelo contrário, as cozinhas da Ásia Oriental e do Sudeste Asiático evitam receitas cujos ingredientes partilham tal afinidade (como é possível observar na figura 4. Nas receitas Norte Americanas quantos mais componentes de sabor dois ingredientes partilharem entre si maior é a probabilidade de aparecerem juntos em receitas. Pelo contrário na cozinha da Ásia Oriental, quantos mais componentes de sabor dois ingredientes partilharem menor é a sua probabilidade de serem utilizados juntos. Contudo o efeito de pairing pode ser explicado apenas por alguns ingredientes, como por exemplo leite, manteiga e ovos para a cozinha Norte Americana ou gengibre, porco e cebola para as cozinhas da Ásia Oriental. Estes ingredientes têm um papel desproporcional em relação aos outros (13 ingredientes chave de receitas Norte Americanas estão presentes em 74.4% de todas as receitas). Em suma, o fitness value de um ingrediente pode variar muito entre cozinhas. A partilha de componentes de sabor representa apenas uma das contribuições para este valor. Em algumas cozinhas esta contribuição é significativa enquanto que em outras não.

### 1.3. A Hipótese de Food-bridging:

A hipótese de food-bridging foi analisada em [2]. Neste artigo é discutido o princípio de food-bridging e é feita uma comparação com o food-pairing partilhando o mesmo dataset e modelos gráficos com [4] enquadrando assim a gastronomia mundial em duas dimensões ao invés de uma só.

Foram observadas quatro grandes classes culinárias: (1) Ásia Oriental, (2) Sudoeste Asiático, (3) cozinhas Ocidentais, (4) América Latina.

Considerando um gráfico ponderado, uma aresta é métrica se o caminho mais curto entre dois ingredientes for igual à aresta por si só (ligação direta). Caso contrário, a aresta é considerada semi-métrica, o que significa que existe pelo menos um caminho alternativo que envolve outros nós (ingredientes) que unem de uma forma mais forte estes dois ingredientes iniciais.

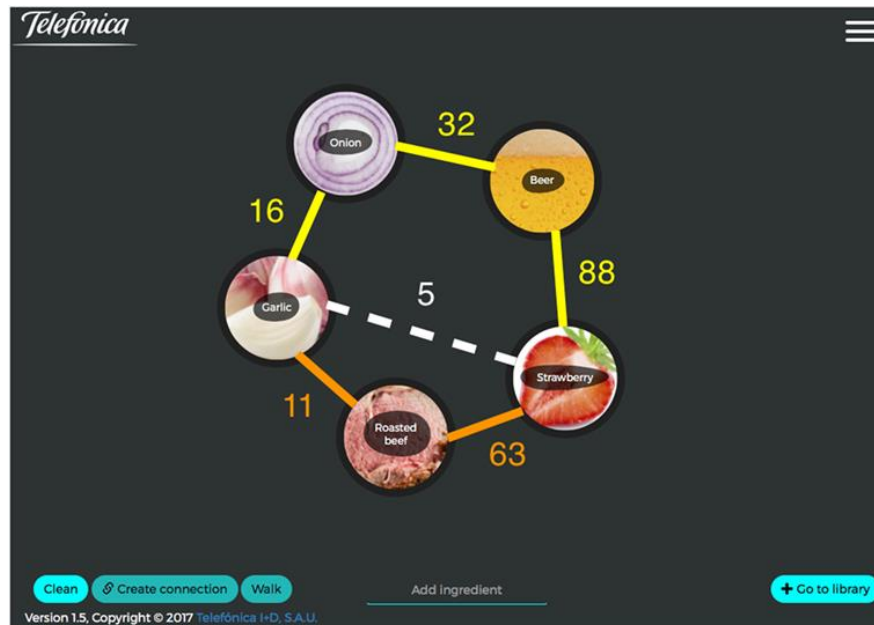
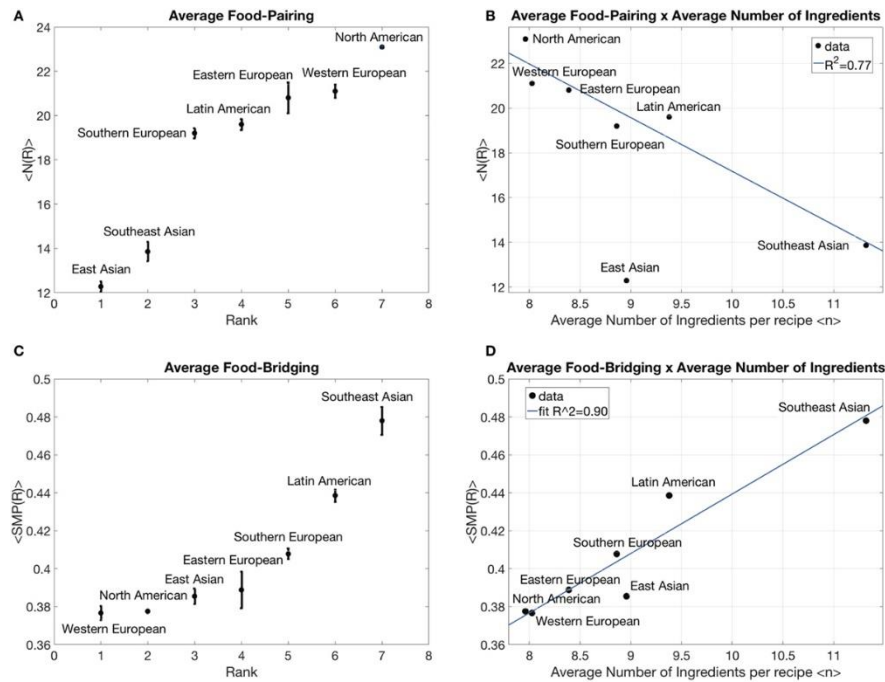


Figura 2 - Arestas semi-métricas e exemplos de caminhos na rede de sabores. A linha branca a tracejado indica que morango e alho partilham 5 sabores. Esta ligação é semi-métrica pois existem dois caminhos alternativos, um a amarelo e outro a laranja. Esta figura é uma captura de ecrã da aplicação desenvolvida por Telefónica I+D Appetit Team e disponível online em <http://appetit.lab.tid.es> [6].



**Figura 3 -  $\langle N(R) \rangle$  é a média do cálculo de food-pairing para todas as receitas,  $\langle SMP(R) \rangle$  é a média do cálculo de food-bridging para todas as receitas, Rank é a classificação do tipo de cozinha por  $\langle N(R) \rangle$  ou  $\langle SMP(R) \rangle$ , e  $\langle n \rangle$  é o número médio de ingredientes por receita de um dado tipo de cozinha.**

Na figura 3 é analisado o food-pairing e o food-bridging. Foi traçada a média destas variáveis para cada uma das sete regiões mundiais distintas em função da sua classificação (Figura 3 A e C) e em função do número de ingredientes (Figura 3 B e D).

#### Food-pairing

- Nas figuras 3 A, B observamos que existem claramente dois grupos distintos no que diz respeito ao food-pairing: as cozinhas de base Ocidental e as cozinhas Asiáticas.
- Esta observação confirma a descoberta feita em [4] de que as cozinhas Asiáticas evitam mais o food-pairing do que as cozinhas baseadas no Ocidente.
- Na figura 3 B observamos uma tendência negativa de food-pairing contra o número médio de ingredientes utilizados numa receita. Nota-se que as cozinhas da Ásia Oriental e do Sudoeste Asiático diferem principalmente no número médio de ingredientes utilizados. A cozinha da Ásia Oriental é um completo outlier nesta tendência, apontando para que a cozinha do Sudoeste Asiático pode divergir da cozinha da Ásia Oriental noutra dimensão.

#### Food-bridging

- Nas figuras 3 C, D observamos que o food-bridging classifica-se como não linear e depende de forma positiva e linear do número médio de ingredientes utilizados nas receitas.

- Observamos que para o caso de food-bridging as cozinhas da Ásia Oriental e do Sudeste Asiático diferem significativamente uma da outra. Cozinhas Ocidentais agrupam-se no extremo inferior enquanto que a cozinha do Sudeste Asiático encontra-se no extremo oposto.
- Observamos também que o food-bridging depende do número de ingredientes utilizados numa receita. Este era um resultado esperado, uma vez que há mais possibilidades de pontes entre ingredientes.

#### Food-pairing e Food-bridging

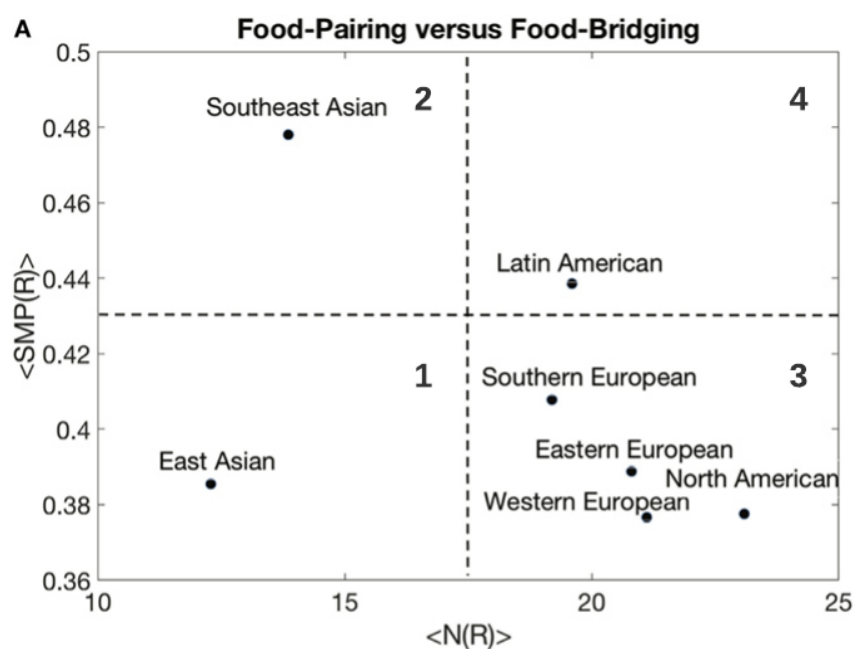


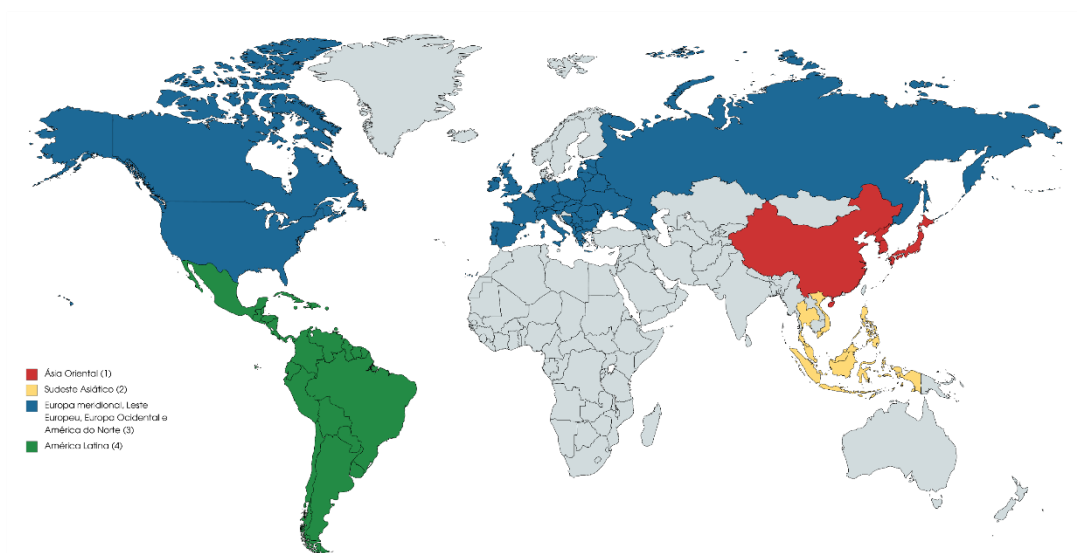
Figura 4 -  $\langle N(R) \rangle$  é a média do cálculo de food-pairing para todas as receitas, e  $\langle SMP(R) \rangle$  é a média do cálculo de food-bridging para todas as receitas.

Na figura 4 observamos a relação entre food-pairing e food-bridging. O grupo foi dividido em quatro regiões não sobrepostas que representam as quatro classes seguintes:

1. Baixo food-pairing + Baixo food-bridging
2. Baixo food-pairing + Alto food-bridging
3. Alto food-pairing + Baixo food-bridging
4. Alto food-pairing + Alto food-bridging

- Na classe 1 os ingredientes de uma receita dependem pouco da co-ocorrência dos seus componentes de sabor, de forma direta ou indireta.
- Na classe 2 combina-se principalmente componentes de sabor indiretamente através de cadeias ou pontes entre ingredientes.
- Na classe 3 os ingredientes combinam os seus componentes de sabor sem muitas cadeias ou pontes indiretas.
- Na classe 4 os ingredientes emparelham e ligam fortemente os seus componentes de sabor.

É ainda possível denotar que as cozinhas representadas acima seguem de alguma forma a sua distribuição geopolítica. Estas cozinhas podem ser influenciadas por limitações geográficas particulares como clima ou recursos, bem como pelo comércio político de bens.



**Figura 5 - Distribuição geográfica das quatro grandes classes culinárias.**

#### 1.4. Digitalização da gastronomia Portuguesa:

Foi ainda feito um estudo para a culinária tradicional portuguesa [5]. Este artigo não é focado nas hipóteses de food-pairing ou food-bridging mas antes apresenta um dataset resultante da abstração de livros de receitas tradicionais da cozinha portuguesa. A análise descritiva feita neste artigo não é no entanto, a sua maior contribuição mas sim a própria criação do dataset [7] que vai permitir dentro da gastronomia computacional fazer análises para a gastronomia Portuguesa. Na criação do dataset apenas foram consideradas

entradas, pratos principais, acompanhamentos e sopas. Sobremesas, bolos, doces, pudins e pastelaria não foram incluídos. As receitas foram caracterizadas por província e ingredientes, independentemente das quantidades ou preparação. Os resultados mostram que as receitas tradicionais portuguesas organizam-se de forma diferente entre as onze províncias consideradas, estabelecendo a base para análises mais detalhadas das 1382 receitas e dos 421 ingredientes presentes no inventário.

As onze províncias consideradas são:



**Figura 6 - Províncias consideradas na construção do dataset.**

As fontes para o dataset foram livros de receitas que tratavam, explicitamente ou visavam apresentar a cozinha tradicional portuguesa. Para serem incluídas no dataset, a origem das receitas tinha de ser conhecida, pelo que, para serem selecionados, os livros de cozinha tinham de permitir que todas ou pelo menos a maioria das receitas fossem atribuídas a qualquer uma das onze províncias e que nenhuma providência fosse omitida, deliberadamente ou não. Foram então utilizados um total de seis livros. Cada linha do dataset apresenta informação para uma única receita em forma colunar. A primeira coluna, intitulada “CODE”, apresenta códigos alfanuméricos para as receitas. A segunda e terceira colunas apresentam o nome da receita em português e inglês, respetivamente. A quarta e quinta colunas (“REF” e “PAGE”) fornecem os números de referências e as páginas onde as receitas completas podem ser encontradas. As restantes 421 colunas exibem os ingredientes individuais, com o valor 1 se o ingrediente estiver presente na receita ou em branco caso contrário.

## Ingredientes

Sal não foi incluído no dataset como ingrediente porque está sempre incluído na receita, em poucos casos onde sal não é mencionado, a sua presença está implícita através de um ou mais ingredientes, geralmente bacalhau. Em relação a ingredientes que são uma mistura de ingredientes, onde se pressupõe que todos sabem como preparar sem necessidade de mais instruções, especificamente acompanhamentos, foi considerado um número mínimo de componentes e estes foram incluídos na receita. Pratos não preparados no momento foram considerados como sendo um único ingrediente, independentemente da sua composição (e.g. enchidos). Em 780 receitas é mencionado pimenta sem especificação do seu tipo, nestes casos é considerado como sendo pimenta branca.

## Receitas

Após a abstração dos seis livros de culinária, o dataset contava com 1644 receitas. Foi realizada uma verificação preventiva de receitas repetidas. Foram feitas comparações aos pares entre cada receita e calculado o seu valor de parença (LI - likeness value):

$$LI = \underbrace{N_{A,B}}_{\text{número de ingredientes presentes em simultâneo nas receitas A e B}} / \underbrace{\max(N_A, N_B)}_{\text{número de ingredientes da receita com maior número de ingredientes}}$$

**Figura 7 - Fórmula para o cálculo do valor de parença (LI - likeness value).**

LI apresenta valores entre [0,1]. Quando LI = 0, as receitas não têm ingredientes em comum. Quando LI = 1, as receitas têm o mesmo número de ingredientes e todos os ingredientes de uma receita ocorrem na outra. No entanto um par de receitas podem ter LI = 1 e ainda serem diferentes, devido à forma como o prato é composto e cozinhado. Por isso foi feita uma verificação adicional à forma de cozinhar para cada par de receitas com LI = 1. O resultado foi a descoberta de 262 receitas que eram verdadeiramente duplicatas, foram assim eliminadas do dataset, que ficou reduzido a 1382 receitas.

## Avaliação do preconceito nos livros de culinária examinados

O dataset foi baseado em seis livros de cozinha de diferentes autores, que tinham objetivos e métodos diferentes de recolha e seleção de receitas. É concebível que os autores e, por conseguinte, os livros e receitas possam ser tendenciosos. O impacto dos supostos preconceitos depende naturalmente da percentagem que cada livro ocupa no total de receitas. Três livros dos seis representam 85% de todas as receitas. Estes foram assim



selecionados para investigações sobre o preconceito que podem apresentar no seu retrato da cozinha tradicional portuguesa. A lógica subjacente ao método utilizado era a de que, se houvesse um preconceito, então a partir da ocorrência de uma combinação de ingredientes numa receita, seria possível prever de onde veio essa receita. Foi adotado um método de classificação não paramétrico com estrutura de árvore. Conclui-se que pode estar presente uma ligeira tendência em torno da pimenta branca e da manteiga.

### Análise descritiva

O Alentejo liderou o ranking do número de receitas (209) presentes no dataset, seguido de Trás-os-Montes e Alto Douro (186) e Estremadura (182). A média das receitas para as onze províncias é de aproximadamente 126. O número médio de ingredientes das 1382 receitas é 9, maior do que o número médio de ingredientes por receita encontrado em inventários mundiais. Uma análise ao coeficiente de assimetria permite concluir que a distribuição do número de ingredientes por província está desequilibrada para a direita, significando que a distribuição apresenta mais peso para valores acima da média. Uma análise individual ao coeficiente de assimetria para cada província permite concluir que só nas duas províncias mais a norte de Portugal, Entre Douro e Minho e Trás-os-Montes e Alto Douro havia indícios de distribuições enviesadas à direita para o número de ingredientes. Uma análise ao valor de parença (LI) permitiu identificar que estes podem estar relacionados com as características das províncias, independentemente do que estas possam ser, tais características apoiam uma hipótese de individualidade. Isto implicaria que quanto mais “única” for uma província (em termos culinários) maior será a semelhança entre as suas receitas e mais pequenos serão estes valores quando comparados com receitas de outra província. Conclui-se que as receitas tradicionais portuguesas organizam-se de forma diferente nas várias províncias. Na caracterização exploratória feita, Entre Douro e Minho e Estremadura ocorrem sempre em conjunto, tal como a Beira Alta e a Beira Baixa. O Alentejo e o Algarve aparecem quase sempre juntos com o primeiro grupo, enquanto que o Ribatejo e os Açores ocorrem juntos com o segundo grupo das Beiras do interior. As duas restantes províncias, Trás-os-Montes e Alto Douro e Beira Litoral aparecem juntamente com a Madeira.

### 1.5. Problema a resolver neste Trabalho Final de Curso:

O dataset gerado e descrito no ponto 1.4. permite fazer um estudo/análise gastronómico para a gastronomia Portuguesa e perante o descrito acima, estão em aberto as seguintes perguntas:

- (1) A gastronomia Portuguesa como se enquadra dentro destas variáveis de food-pairing versus food-bridging?
- (2) Como variam estas variáveis pelas regiões de Portugal?



## 1.6. Materiais e Métodos

### 1.6.1. Materiais

Os materiais utilizados neste trabalho foram: (1) o dataset português [5,7] e a Flavor Network [4]. Ao dataset português foram aplicadas algumas transformações:

1. O primeiro passo foi realizar a correspondência entre os ingredientes presentes na Flavor Network com os ingredientes presentes no dataset português. Este passo é crucial pois permite obter os componentes de sabor correspondentes a cada ingrediente do dataset português. Este passo envolve manualmente percorrer cada ingrediente do dataset português e encontrar a correspondência na Flavor Network, algo que nem sempre é trivial dada a abundância de designações que o mesmo ingrediente pode ter.
  - a. Existem um total de 421 ingredientes no dataset português
  - b. Foi encontrada correspondência para 263 destes ingredientes.
2. O segundo passo foi eliminar do dataset portuguesas receitas que tivessem pelo menos um ingrediente para o qual não foi encontrada correspondência.
  - a. Existem um total de 1382 receitas no dataset português
  - b. Após estes processos sobraram 495 receitas

Região (Code)	Percentagem de receitas perdidas
Açores (AÇOR)	66.7
Alentejo (ALEN)	68.9
Algarve (ALGA)	49
Beira-Alta (BALT)	60.8
Beira-Baixa (BBAI)	73
Beira-Litoral (BLIT)	65.1
Entre Douro e Minho (EDMI)	64.4
Estremadura (ESTR)	60.4
Madeira (MADE)	51.4
Ribatejo (RIBA)	65.9
Trás-os-Montes e Alto Douro (TMAD)	74.2

3. O terceiro passo foi listar o número de componentes de sabor que cada par de ingredientes partilha. Os componentes de sabor de cada ingrediente fazem parte da Flavor Network. Em primeiro lugar foi feita a listagem para cada ingrediente e os componentes de sabor que o constituem e de seguida foi feita a comparação entre cada ingrediente e os demais para obter o número de componentes de sabor que partilham.

### 1.6.2. Métodos

Os métodos utilizados neste trabalho são baseados no trabalho de Ah et al. [4] e Simas et al. [2].

#### 1.6.2.1. Food-Pairing e Food-Bridging

O valor de food-pairing de uma receita é a média do número de componentes de sabor dessa receita e é calculado utilizando a seguinte equação:

$$N_S(R) = \frac{2}{nR(nR - 1)} \sum_{i,j \in R, i \neq j} |C_i \cap C_j|$$

onde  $C_k$  corresponde ao peso da aresta entre o par de ingredientes na Flavor Network e  $nR$  é o número de ingredientes na receita  $R$ . Cada receita define um subgrafo na Flavor Network e  $N_S(R)$  corresponde à média de todas as aresta nesse subgrafo.

O valor de food-bridging é calculado utilizando a seguinte equação:

$$SMP = \frac{\sum_{i,j} \delta(s_{i,j} > 1 \wedge s_{i,j} < +\infty)}{\sum_{i,j} \delta(s_{i,j} \geq 1 \wedge s_{i,j} < +\infty)}$$

onde  $s_{i,j}$  é o rácio de semi-metricidade entre os ingredientes  $i$  e  $j$  na Flavor Network. O denominador desta equação conta todas arestas na rede e o numerado conta apenas as arestas semi-métricas.

#### 1.6.2.2. Frequência de utilização de cada ingrediente em receitas portuguesas

Utilizando o dataset português é possível calcular a frequência de utilização de cada ingrediente para todas as receitas e para cada região. Cada linha do dataset é uma receita e cada coluna é um ingrediente, basta então verificar para cada linha quais os ingredientes presentes e calcular a percentagem.

#### 1.6.2.3. Combinações mais utilizadas

Para este passo é em primeiro lugar necessário gerar a matriz de cada recita. Essa matriz é constituída pelo número de componentes de sabor que cada ingrediente partilha entre si. Em seguida é necessário passar esta matriz para o espaço das distâncias. Por fim é calculada a semi-metricidade entre cada par de ingredientes.

## **2 Viabilidade e Pertinência**

Com a finalização deste trabalho, foi possível não só enquadrar a gastronomia portuguesa de acordo com o food-pairing e food-bridging como também calcular algumas métricas interessantes que nos providenciam de um conhecimento mais aprofundado sobre a nossa gastronomia, como por exemplo a frequência de utilização de cada ingrediente para todas as receitas ou para cada região individual. Com os resultados aqui obtidos está em fase de preparação um manuscrito (paper) para submeter a uma revista científica a selecionar apos a escrita do manuscrito.

### **3 Solução Desenvolvida**

O resultado final deste trabalho é um estudo de âmbito científico que nos permite ficar a conhecer um pouco melhor a gastronomia portuguesa e como esta se enquadra dentro das hipóteses de food-pairing e food-bridging. O manuscrito mencionado anteriormente vai assim permitir um incremento científico ao enquadramento da gastronomia portuguesa no mundo.

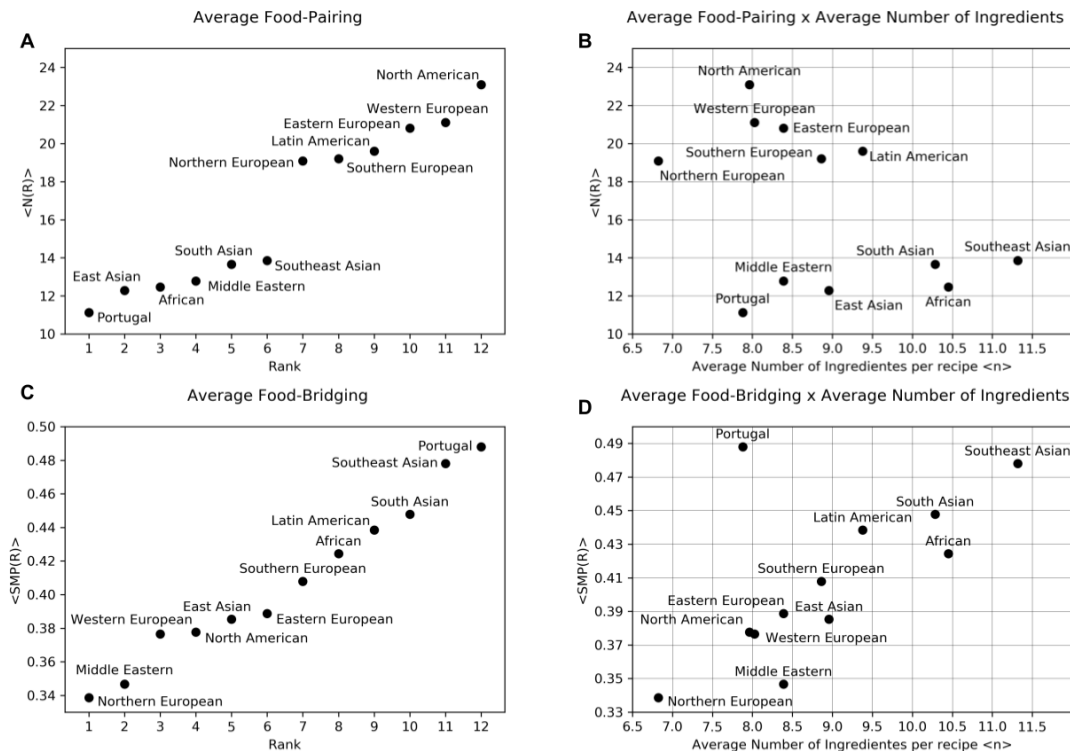
O resultado sofre de algumas limitações nomeadamente na quantidade de receitas utilizadas bem como no número de ingredientes. Dos 421 ingredientes e das 1382 receitas presentes no dataset português [5] apenas foram utilizados 263 ingredientes e 465 receitas.

## **4 Benchmarking**

O artigo [2] e [4] devido a sua similaridade com este TFC permite fazer uma análise comparativa com os resultados aqui obtidos. Estes artigos utilizam métricas semelhantes aquelas utilizadas neste TFC mas para explorar e enquadrar sete regiões mundiais nas hipóteses de food-pairing e food-bridging. Utilizando assim o artigo [2] como base foi possível comparar Portugal com os resultados já obtidos para as restantes regiões e também utilizar os seus métodos para realizar um estudo para cada região de Portugal. Em seguida são apresentados os resultados com base nos materiais e métodos.

## 5 Resultados

### 5.1 Comparar Portugal com as restantes regiões do mundo



**Figura 8** -  $\langle N(R) \rangle$  é a média do cálculo de food-pairing para todas as receitas,  $\langle SMP(R) \rangle$  é a média do cálculo de food-bridging para todas as receitas, Rank é a classificação do tipo de cozinha por  $\langle N(R) \rangle$  ou  $\langle SMP(R) \rangle$ , e  $\langle n \rangle$  é o número médio de ingredientes por receita de um dado tipo de cozinha.

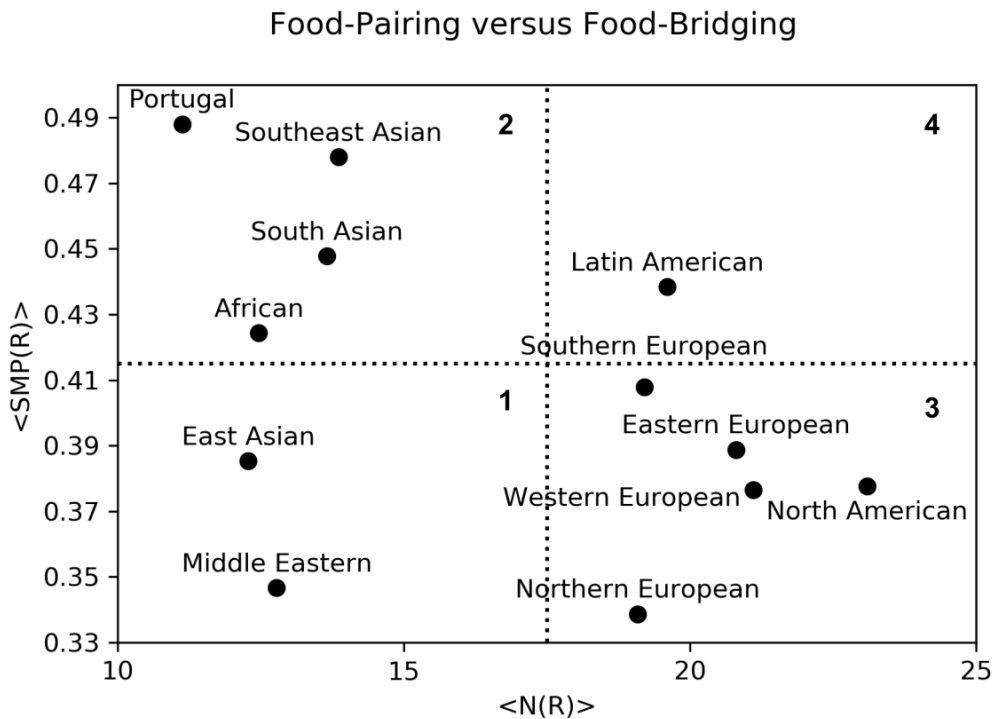
Na figura 8 é analisado o food-pairing e o food-bridging. Foi traçada a média destas variáveis para cada uma das doze regiões mundiais distintas em função da sua classificação (Figura 8 A e C) e em função do número de ingredientes (Figura 8 B e D).

#### Food-pairing

- Já tínhamos identificado anteriormente dois grupos distintos em relação às figuras A e B. Estes grupos dizem respeito às cozinhas de base Ocidental e às cozinhas de base Oriental. Porém com a adição de Portugal reparamos que a sua posição não é onde inicialmente seria de esperar dada a sua localização geográfica. Portugal encontra-se com o grupo de base Oriental, sendo a região do mundo abrangida por este estudo que mais evita o food-pairing.

Food-bridging

- Já tínhamos identificado que o food-bridging classifica-se como não linear e dependente de forma positiva e linear do número médio de ingredientes utilizados nas receitas. Porém Portugal é uma exceção sendo claramente um outlier nesta tendência como podemos observar na figura D. Na figura C vemos também que de todas as regiões abrangidas neste estudo, Portugal é a região que faz mais uso do food-bridging, apontando assim para a grande riqueza gastronómica do nosso país.



**Figura 9 -  $\langle N(R) \rangle$  é a média do cálculo de food-pairing para todas as receitas, e  $\langle SMP(R) \rangle$  é a média do cálculo de food-bridging para todas as receitas.**

Na figura 9 observamos a relação entre food-pairing e food-bridging. A figura foi dividida em quatro regiões não sobrepostas que representam as quatro classes seguintes:

1. Baixo food-pairing + Baixo food-bridging
2. Baixo food-pairing + Alto food-bridging
3. Alto food-pairing + Baixo food-bridging
4. Alto food-pairing + Alto food-bridging

Podemos observar que Portugal encontra-se na segunda região, isto significa que se combinam principalmente componentes de sabor indiretamente através de cadeias ou pontes entre ingredientes.

## 5.2 Comparar as diversas regiões de Portugal

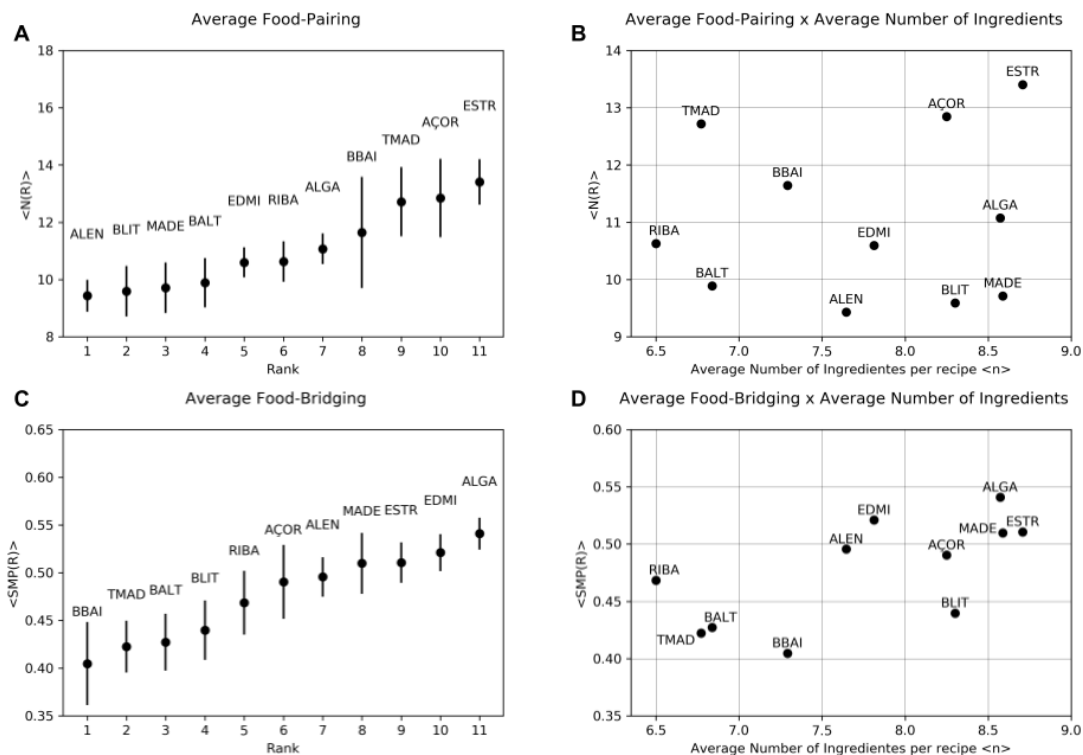


Figura 10 -  $\langle N(R) \rangle$  é a média do cálculo de food-pairing para todas as receitas,  $\langle SMP(R) \rangle$  é a média do cálculo de food-bridging para todas as receitas, Rank é a classificação do tipo de cozinha por  $\langle N(R) \rangle$  ou  $\langle SMP(R) \rangle$ , e  $\langle n \rangle$  é o número médio de ingredientes por receita de um dado tipo de cozinha.

Na figura 10 é analisado o food-pairing e o food-bridging. Foi traçada a média destas variáveis para cada uma das onze regiões distintas em função da sua classificação (Figura A e C) e em função do número de ingredientes (Figura B e D).

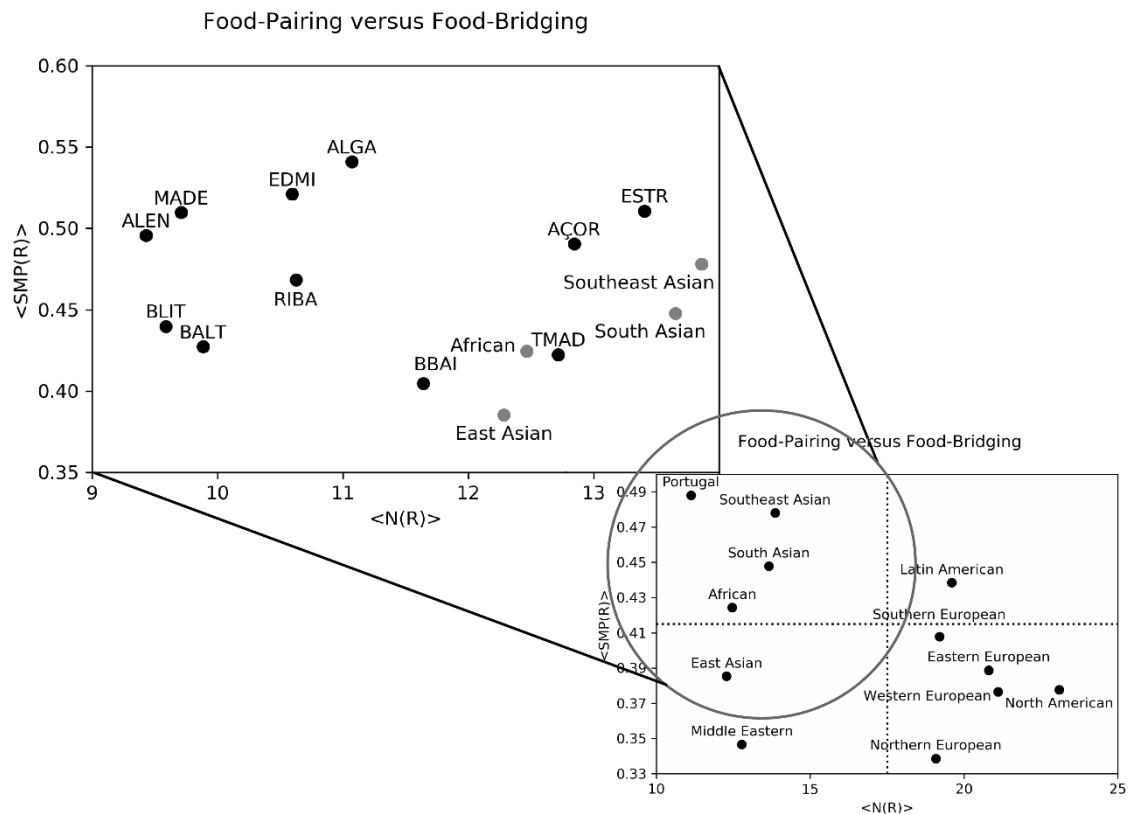
### Food-pairing

- Na figura A podemos observar que o erro padrão das regiões da Beira-Baixa (BBAI), Trás-os-Montes e Alto Douro (TMAD) e Açores (AÇOR) é bastante elevado. Isto significa que tanto existem receitas que fazem pouco uso de food-pairing como existem receitas onde a hipótese demonstra-se bastante utilizada.

### Food-bridging

- Na figura D podemos identificar alguns clusters como é o caso da região de Trás-os-Montes e Alto Douro (TMAD) e Beira-Alta (BALT) ou Madeira (MADE) e Estremadura (ESTR), ambos extremamente similares no que toca ao food-bridging e ao número de ingredientes.





**Figura 11** -  $\langle N(R) \rangle$  é a média do cálculo de food-pairing para todas as receitas, e  $\langle SMP(R) \rangle$  é a média do cálculo de food-bridging para todas as receitas.

Na figura 11 observamos a relação entre food-pairing e food-bridging. Foi feita uma sobreposição com as regiões mundiais que apresentam valores similares aos das regiões portuguesas.

5.3 Frequência de utilização de cada ingrediente em receitas portuguesas

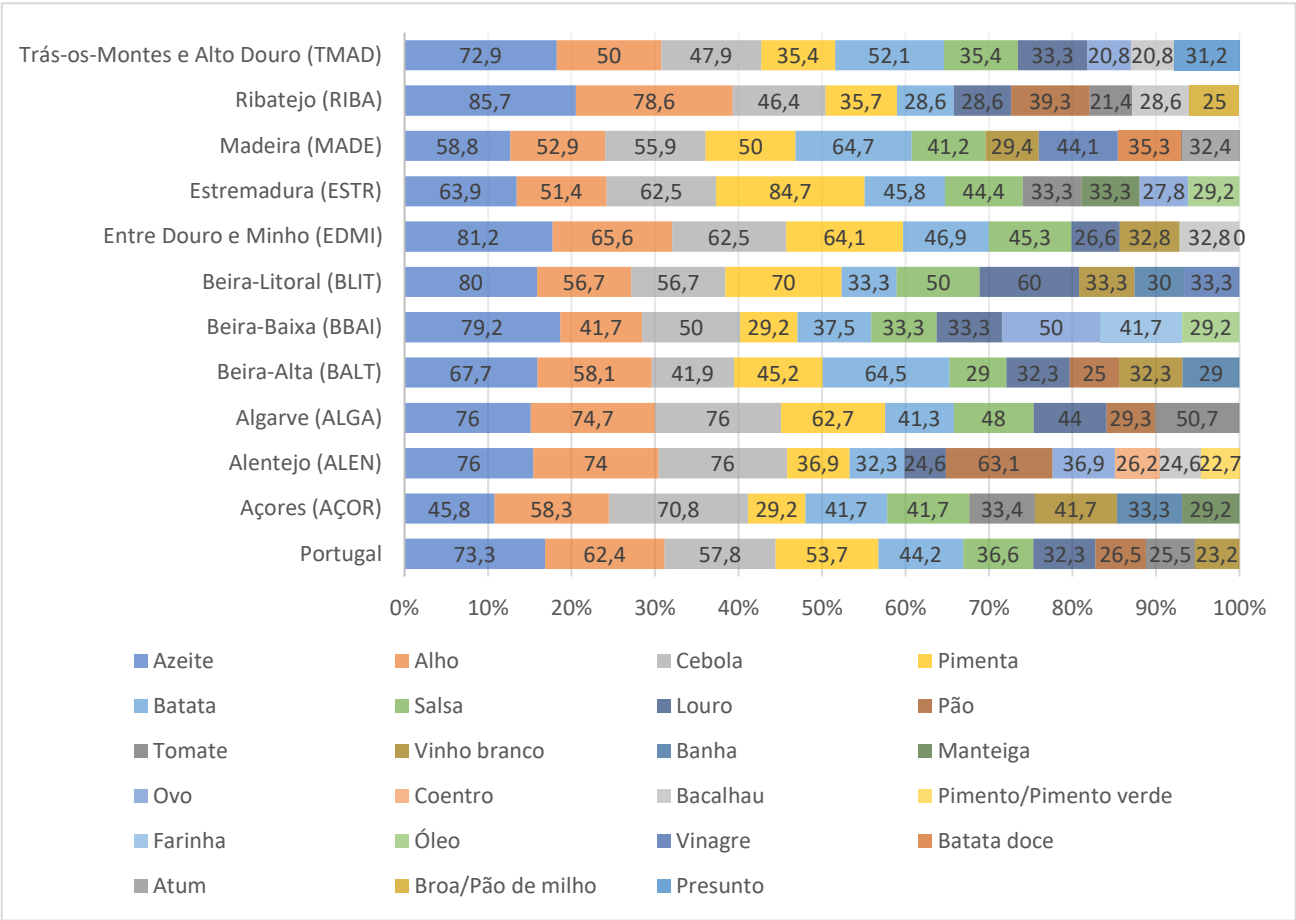


Figura 12 - Ingredientes mais utilizados por região. Esta figura demonstra por percentagem a utilização dos 10 ingredientes mais frequentes em receitas Portuguesas como um todo e por região.

## 5.4 Combinações de ingredientes mais utilizadas

Tabela 1 - As 10 combinações mais utilizadas - Portugal

Ingrediente 1	Ligação 1	Ingrediente 2	Frequência
Alho	Salsa	Pimenta	20
Alho	Salsa	Cebola	16
Azeite	Tomate	Broa/Pão	14
Alho	Salsa	Louro	14
Azeite	Salsa	Cebola	13
Azeite	Alho	Broa/Pão	12
Azeite	Batata	Bacalhau	11
Alho	Salsa	Batata	10
Alho	Pimenta	Cebola	10
Alho	Pimenta	Louro	10

Tabela 2 - As 5 combinações mais utilizadas – Açores (AÇOR)

Ingrediente 1	Ligação 1	Ingrediente 2	Frequência
Azeite	Tomate	Vinho branco	3
Funcho	Banha	Toucinho	3
Vinagre	Tomate	Vinho branco	2
Cebola	Batata	Salsa	2
Peixes vários <sup>1</sup>	Tomate	Vinho branco	13

<sup>1</sup> Não foi possível estabelecer com certeza o peixe em específico devido a falta de correspondências entre os peixes presentes em receitas portuguesas e aqueles presentes na Flavor Network.

**Tabela 3 - As 5 combinações mais utilizadas – Alentejo (ALEN)**

<b>Ingrediente 1</b>	<b>Ligação 1</b>	<b>Ingrediente 2</b>	<b>Frequência</b>
Azeite	Alho	Broa/Pão	7
Ovo	Alho	Broa/Pão	5
Azeite	Tomate	Broa/Pão	5
Azeite	Alho	Queijo fresco	5
Alho	Tomate	Azeite	4

**Tabela 4 - As 5 combinações mais utilizadas – Algarve (ALGA)**

<b>Ingrediente 1</b>	<b>Ligação 1</b>	<b>Ingrediente 2</b>	<b>Frequência</b>
Alho	Tomate	Pimento verde	6
Azeite	Tomate	Broa/Pão	5
Broa/Pão	Tomate	Toucinho	5
Louro	Tomate	Broa/Pão	5
Alho	Salsa	Cebola	4

**Tabela 5 - As 5 combinações mais utilizadas – Beira-Alta (BALT)**

<b>Ingrediente 1</b>	<b>Ligação 1</b>	<b>Ingrediente 2</b>	<b>Frequência</b>
Cebola	Broa/Pão	Vinho branco	3
Louro	Tomate	Broa/Pão	2
Azeite	Vinho branco	Louro	2
Alho	Azeite	Pimenta	2
Batata	Salsa	Óleo	2

Tabela 6 - As 5 combinações mais utilizadas – Beira-Baixa (BBAI)

Ingrediente 1	Ligação 1	Ligação 2	Ingrediente 2	Frequência
Banha	Vinho branco	-	Salsa	2
Banha	Vinho branco	-	Batata	2
Alho	Salsa	Vinho branco	Banha	2
Alho	Salsa	-	Batata	2
Alho	Salsa	-	Vinho branco	2

Tabela 7 - As 5 combinações mais utilizadas – Beira-Litoral (BLIT)

Ingrediente 1	Ligação 1	Ingrediente 2	Frequência
Alho	Salsa	Pimenta	3
Alho	Vinagra	Louro	3
Azeite	Toucinho	Louro	2
Alho	Bacalhau	Batata	2
Cebola	Salsa	Farinha	2

Tabela 8 - As 5 combinações mais utilizadas – Entre Douro e Minho (EDMI)

Ingrediente 1	Ligação 1	Ingrediente 2	Frequência
Alho	Salsa	Cebola	5
Alho	Salsa	Pimenta	4
Azeite	Batata	Bacalhau	4
Bacalhau	Batata	Pimenta	4
Azeite	Santola/Sapateira	Pimenta	3

Tabela 9 - As 5 combinações mais utilizadas – Estremadura (ESTR)

Ingrediente 1	Ligação 1	Ingrediente 2	Frequência
Cebola	Pimenta	Tomate	5
Alho	Pimenta	Tomate	4
Alho	Salsa	Pimenta	4
Azeite	Pimenta	Cebola	4
Manteiga	Batata	Pimenta	4

Tabela 10 - As 5 combinações mais utilizadas – Madeira (MADE)

Ingrediente 1	Ligação 1	Ingrediente 2	Frequência
Azeite	Salsa	Cebola	6
Atum	Vinagre	Azeite	5
Alho	Batata	Azeite	5
Alho	Batata	Vinagre	5
Azeite	Salsa	Batata	4

Tabela 11 - As 5 combinações mais utilizadas – Ribatejo (RIBA)

Ingrediente 1	Ligação 1	Ingrediente 2	Frequência
Azeite	Alho	Broa/Pão	4
Azeite	Alho	Bacalhau	3
Bacalhau	Alho	Broa/Pão	3
Alho	Broa/Pão	Tomate	3
Broa/Pão	Feijão	Couve	2

**Tabela 12 - As 5 combinações mais utilizadas – Trás-os-Montes e Alto Douro (TMAD)**

<b>Ingrediente 1</b>	<b>Ligação 1</b>	<b>Ligação 2</b>	<b>Ingrediente 2</b>	<b>Frequência</b>
Alho	Salsa	-	Pimenta	2
Louro	Pimenta	Cebola	Salsa	2
Cebola	Pimenta	-	Louro	2
Alho	Bacalhau	-	Pão de centeio	2
Alho	Bacalhau	-	Batata	2

## **6 Método e planeamento**

No início do ano letivo definimos um calendário para a realização deste TFC. Esse calendário consistia em duas grandes tarefas: (1) Aferir como se enquadra a gastronomia Portuguesa dentro das variáveis de food-pairing versus food-bridging e (2) - Aferir como variam estas variáveis pelas regiões de Portugal.

Contudo houve uma fase anterior, em que o aluno ficou familiarizado com o estado da arte da gastronomia computacional, nomeadamente através da leitura dos papers [2,4,8].

Em seguida elaboramos um plano de trabalho que seguimos até ao fim. Agendamos reuniões semanais onde uma nova tarefa era atribuída ao aluno. Essa tarefa era realizada e os resultados partilhados na reunião da semana seguinte onde por sua vez uma nova tarefa era atribuída. Este método de trabalho permitiu que cada iteração fosse realizada e validada no espaço de uma semana ajudando assim à melhor compreensão de conceitos como food-pairing e food-bridging totalmente novos para o aluno bem como a um melhor entendimento de matérias já lecionadas como teoria dos grafos e algoritmos de busca.



## 7 Conclusão

Analisando os resultados obtidos neste trabalho podemos formular algumas hipóteses para os explicar. Desde já, e sem dúvida um dos resultados mais fascinantes deste trabalho é o valor de food-bridging apresentado por Portugal. Apresenta um valor mais elevado do que qualquer outra região estudada anteriormente (figura 8 C). Podemos aqui especular que tal fenómeno se deva aos Descobrimentos, onde os portugueses foram coletando ingredientes e saberes culinários que em última instância tornam-se parte da nossa coletânea gastronómica. Também não deixa de ser interessante que em relação do food-pairing (figura 8 A) Portugal encontra-se no extremo oposto aos restantes países europeus. Talvez seja essa diferença que torna a culinária Portuguesa fascinante aos olhos de muitos turistas.

Em relação as diferentes regiões Portuguesas, explicações como as dadas em cima já envolvem um conhecimento mais aprofundado sobre cada região e especialmente sobre como a gastronomia difere entre elas. Contudo existem algumas curiosidades interessantes a apontar. Como por exemplo a similaridade entre a gastronomia Africana e a de Trás-os-Montes e Alto Douro (figura 11) ou a presença de combinações peculiares/características em algumas regiões, como funcho, banha e toucinho nos Açores (Tabela 2); azeite, alho e queijo fresco no Alentejo (tabela 3); alho, tomate e pimento verde no Algarve (tabela 4); alho, bacalhau e batata na Beira-Litoral (tabela 7) ou broa/pão, feijão e couve no Ribatejo (tabela 11).



## **Bibliografia**

- [1] Blumenthal, H. The big fat duck cookbook (Bloomsbury, London, 2008).
- [2] Simas T, Ficek M, Diaz-Guilera A, Obrador P and Rodriguez PR (2017) Food-Bridging: A New Network Construction to Unveil the Principles of Cooking. *Front. ICT* 4:14. doi: 10.3389/fict.2017.00014
- [3] Burdock, G.A. (2010). *Fenaroli's Handbook of Flavor Ingredients*. 6th edition, CRC Press.
- [4] Ahn, Y.-Y., Ahnert, S. E., Bagrow, J. P., and Barabasi, A.-L. (2011). Flavor network and the principles of food pairing. *Sci. Rep.* 1, 196. doi:10.1038/srep00196
- [5] Soveral Dias A., Silva Dias L. (2018) A Data Set of Portuguese Traditional Recipes Based on Published Cookery Books
- [6] <http://appetit.lab.tid.es>
- [7] [www.mdpi.com/2306-5729/3/1/9/s1](http://www.mdpi.com/2306-5729/3/1/9/s1)
- [8] Weiqing Min, Shuqiang Jiang, Linhu Liu, Yong Rui, and Ramesh Jain. 2019. A Survey on Food Computing.
- [9] [https://github.com/a21703925/DEISI203\\_TFC](https://github.com/a21703925/DEISI203_TFC)

## **Glossário**

LEI      Licenciatura em Engenharia Informática

TFC      Trabalho Final de Curso

**Food-pairing**    Hipótese que defende que se dois ingredientes partilharem um maior número de componentes de sabor, estes têm maior probabilidade de combinarem melhor entre si do que ingredientes que não partilhem estas afinidades químicas

**Food-bridging**    Hipótese que defende que se dois ingredientes não partilharem uma afinidade molecular ou empírica como descrito no food-pairing, eles podem tornar-se afins através de uma cadeia de afinidades que envolve outros ingredientes que fazem a ponte entre os ingredientes de pouca ou nenhuma afinidade.